Supplementary Information for

Common genetic variants in the *CLDN2* and *PRSS1-PRSS2* loci alter risk for alcohol-related and sporadic pancreatitis

David C. Whitcomb, Jessica LaRusch, Alyssa M. Krasinskas, Lambertus Klei, Jill P. Smith, Randall E. Brand, John P. Neoptolemos, Markus M. Lerch, Matt Tector, Bimaljit S. Sandhu, Nalini M. Guda, Lidiya Orlichenko, Alzheimer's Disease Genetics Consortium, Samer Alkaade, Stephen T. Amann, Michelle A. Anderson, John Baillie, Peter A. Banks, Darwin Conwell, Gregory A. Coté, Peter B. Cotton, James DiSario, Lindsay A. Farrer, Chris E. Forsmark, Marianne Johnstone, Timothy B. Gardner, Andres Gelrud, William Greenhalf, Jonathan L. Haines, Douglas J. Hartman, Robert A. Hawes, Christopher Lawrence, Michele Lewis, Julia Mayerle, Richard Mayeux, Nadine M. Melhem, Mary E. Money, Thiruvengadam Muniraj, Georgios I. Papachristou, Margaret A. Pericak-Vance, Joseph Romagnuolo, Gerard D. Schellenberg, Stuart Sherman, Peter Simon, Vijay K. Singh, Adam Slivka, Donna Stolz, Robert Sutton, Frank Ulrich Weiss, C. Mel Wilcox, Narcis Octavian Zarnescu, Stephen R. Wisniewski, Michael R. O'Connell, Michelle L. Kienholz, Kathryn Roeder, M. Michael Barmada, Dhiraj Yadav, Bernie Devlin

Supplementary Note: Protein-protein and system-based interactions among genes with variants associated with recurrent acute and chronic pancreatitis.

The primary protein-protein interactions related to pancreatic disease involve trypsin (*PRSS1*) interacting with chymotrypsin C (*CTRC*) and pancreatic secretory trypsin inhibitor (*SPINK1*). When activated by trypsin, Chymotrypsin C degrades trypsin in low calcium environments (e.g., within the pancreatic acinar cell), thereby protecting the pancreas from premature trypsin activation[1,2]. Pancreatic secretory trypsin inhibitor is an acute phase protein (markedly up regulated by inflammation) and is a suicide inhibitor of trypsin[3,4].

Other protein-protein interactions are higher order, and their effects are indirect (See Supplement Figure 1). For example, the pancreatic duct cells protect the pancreas from trypsin injury by secreting a bicarbonate-rich fluid using a CFTR-dependent mechanism to flush active trypsin out of the pancreas. Mild-variable mutations in the cystic fibrosis gene (i.e., *CFTR*) result in failed bicarbonate secretion, which results in a pH within the duct lumen that is favorable for trypsin activation. Recurrent trypsin activation results in injury and inflammation, which is countered by increased *SPINK1* expression and the protection thus conferred. Genetic variants in *SPINK1* limit the protection from premature trypsin activation, resulting in ongoing injury and the development of chronic pancreatitis over time. Thus, mild-variable *CFTR* variants are associated with chronic pancreatitis when they occur together with *SPINK1* mutations, and vice versa, even though the proteins do not interact directly[5]. Similarly, the calcium sensing receptor (*CASR*) is located on the luminal side of the duct and monitors calcium concentrations[6]. When activated, CaSR, a G-coupled receptor, initiates a second messenger signal that opens CFTR to flush the duct and reduce calcium concentrations to limit trypsin-activation[1].

In the current paper, we hypothesize that a relative reduction of *PRSS1* expression reduces the amount of trypsinogen in the pancreas and in turn reduces risk of activation, as is observed in genetic mouse models[7]. The other protein, claudin-2, is expected to act in parallel with CFTR in the duct. CFTR facilitates bicarbonate (HCO3-) secretion into the duct lumen through the duct cell; claudin-2 facilitates transport of sodium (Na+) and water between the ductal epithelial cells (paracellular flow)[8,9]. The sodium and bicarbonate meet within the duct to produce pancreatic juice. Thus, claudin-2 and CFTR do not directly interact, but their functions are intimately linked. Of note, claudin-2 is also up-regulated during inflammation[10-12], and we hypothesize that it is the abnormal regulation and localization of claudin-2 in response to injury and inflammation in patients with the high-risk *CLDN2*-locus genotype that leads to chronic pancreatitis. While plausible, this hypothesis will require further study to demonstrate that the mechanism is linked to the well-documented role of macrophages[13]. We identified claudin-2-positive macrophages, as well as claudin-2-positive acinar cells, in patients with the high-risk *CLDN2* genotype (Figure 2 C&F) and speculate is the presence of a pathologic interaction, as proposed by others[14].

Supplementary Table 1. Demographic characteristics of pancreatitis patients and controls (see footnote 1 for details of recruitment).

Chronic Pancreatitis

				Age	
Study	**n**	**male**	**female**	**mean**	**sd**
ALL	930	501	429	48.6	17.0
German	150	84	66	39.9	15.9
HP	53	19	34	35.9	19.4
Milwaukee	45	30	15	61.8	15.4
NAPS-CV	275	150	125	51.7	15.3
NAPS2	406	217	189	49.7	16.2
PAGER	1	1		54.0	

Recurrent Acute Pancreatitis

				Age	
Study	**n**	**male**	**female**	**mean**	**sd**
ALL	576	247	329	44.4	16.5
German	91	38	53	40.2	16.9
HP	30	11	19	32.8	20.6
Milwaukee	25	4	21	52.3	15.9
NAPS-CV	53	29	24	43.7	14.6
NAPS2	373	163	210	45.9	15.8
PROOF	2	1	1	33.0	4.2
SAPS	2	1	1	62.5	7.8

Control

				Age	
Study	**n**	**male**	**female**	**mean**	**sd**
ALL	538	213	325	55.1	15.7
NAPS-CV	10	4	6	46.1	11.4
NAPS2	386	149	237	53.9	14.9
PAGER	10	2	8	66.2	13.7
Somalogic	132	58	74	58.4	17.3

[1]The study included case and control samples of European ancestry, which was also validated genetically (Supplementary Figs. 1 & 2). The North American Pancreatitis Study 2 (NAPS2) prospectively ascertained and phenotyped 540 chronic pancreatitis (CP) patients, 460 recurrent acute pancreatitis (RAP) patients, and 695 control subjects from 20 US expert clinical centers between August 2000 and September 2006 as described[15]. The NAPS2 continuation and validation study (NAPS2-CV) ascertained an additional 516 CP patients through March 2012.

Parallel studies at the University of Pittsburgh that used the same NAPS2 case report forms and methods included the Severity of Acute Pancreatitis Study / Pancreatitis-associated Risk of Organ Failure (SAPS/PROOF) study [16] (Papachristou, PI), which ascertained RAP patients between June 2003 until March 2012; and the Pancreatic Adenocarcinoma Gene-Environment Risk (PAGER) study [17] (Brand, PI), ascertained controls between February 2002 and March 2012. Only Caucasian cases and controls that met NAPS2 criteria for RAP, CP, or control were included. Cases from the Hereditary Pancreatitis (HP) study [18] (Whitcomb, PI) were primarily probands of small familial pancreatitis families or spouses of affected individuals as controls. DNA samples from Milwaukee were obtained from Aurora Health Care system, Saint Luke's Hospital, Milwaukee, WI, as part of the Open-Source Robotic Biorepository & Information Technology (ORBIT) program (Tector, Director) using de-identified DNA samples and case forms completed by an honest broker using electronic medical records. Liverpool samples were collected as clinical referrals (Neoptolemos, PI), while the German samples were prospectively collected from clinical visits and referrals centered in Munster, DE and Greifswald, DE (Lerch, PI). Additional phenotyped Caucasian control DNA samples were provided in collaboration with SomaLogic, Inc (Boulder CO). Alzheimer Disease Genetics Consortium (ADGC) samples consisted of cases and controls genotyped on the Illumina HumanOmniExpress Beadchips and judged to be of European ancestry by genetic analysis[19]. The NeuroGenetics Research Consortium (NGRC[20]) collected a set of 2000 cases and 2000 controls for the Genome-Wide Association Study of Parkinson Disease: Genes and Environment. These samples, which were used as controls for the Stage 2 GWAS, were genotyped on the Illumina 1M. Data from these samples are obtained from dbGaP Study Accession: phs000196.v2.p1 (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000196.v2.p1), where a detailed study description can be found. These subjects are consented for general use genetics studies.

Supplementary Table 2. SNPs from Stage 1, Stage 2, and the joint analysis (SNPs passing quality control, have $p < 5 \times 10^{-7}$ for the Stage 1 or Stage 2 or combined analysis based on results from Plink).

					CP + RAP		CP			CP + RAP			CP + RAP		
					Allele Frequency (A1)		Stage 1			Stage 2			Combined		
CHR	SNP	BP	A1*	A2	cases	controls	OR	se(OR)	P	OR	se(OR)	P	OR	se(OR)	P
7	rs10273639	142456928	T	C	0.350	0.424	0.712	0.044	3.0×10^{-8}	0.748	0.039	7.5×10^{-8}	0.734	0.029	2.0×10^{-14}
8	rs11988997	119766194	T	C	0.094	0.071	1.086	0.120	4.6×10^{-1}	1.593	0.095	1.1×10^{-7}	1.359	0.074	6.1×10^{-6}
10	rs2995271	30519832	C	T	0.196	0.238	0.780	0.058	7.8×10^{-4}	0.794	0.050	3.2×10^{-4}	0.789	0.038	8.2×10^{-7}
X	rs379742	105493919	T	C	0.250	0.200	1.392	0.077	1.8×10^{-9}	1.076	0.070	1.5×10^{-1}	1.203	0.052	5.9×10^{-7}
X	rs4409525	106140325	A	G	0.372	0.281	1.486	0.075	4.1×10^{-15}	1.195	0.066	6.4×10^{-5}	1.310	0.049	4.5×10^{-16}
X	rs7057398	106144529	C	T	0.374	0.281	1.493	0.075	1.4×10^{-15}	1.210	0.066	1.8×10^{-5}	1.321	0.049	4.6×10^{-17}
X	rs12008279	106160702	G	A	0.509	0.437	1.338	0.065	2.7×10^{-9}	1.126	0.056	4.9×10^{-3}	1.212	0.043	1.6×10^{-9}
X	rs5917027	106162634	C	T	0.515	0.438	1.330	0.065	5.6×10^{-9}	1.158	0.056	4.9×10^{-4}	1.229	0.042	9.7×10^{-11}
X	rs12014762	106183670	T	C	0.279	0.202	1.523	0.081	2.2×10^{-15}	1.182	0.075	6.5×10^{-4}	1.318	0.055	1.5×10^{-14}
X	rs6622126	106200202	A	G	0.510	0.435	1.329	0.065	6.1×10^{-9}	1.154	0.056	6.8×10^{-4}	1.225	0.042	1.9×10^{-10}
X	rs12688220	106244767	T	C	0.367	0.261	1.612	0.081	2.4×10^{-21}	1.238	0.073	2.3×10^{-6}	1.385	0.054	2.3×10^{-22}

*A1 is the allele counted for purposes of computing odds ratio and associated statistics. The model used here includes covariates to control for the two leading eigenvectors for ancestry, as was done in the Plink analyses, but differs in its treatment of the minor allele count for the *CLDN2* locus, which resides on the X chromosome. In this case, Plink encodes the count of minor alleles in males as 0 and 1 and includes a sex effect; instead, following Clayton (2009; PMID: 19939292), we model the male genotypes as 0 and 2 and do not include the sex effect, because this is a more powerful approach. Alleles given are refSNP alleles according to dbSNP, which are not necessarily the alleles designated in the Illumina map.

Supplementary Table 3. Odds ratios (OR), the standard error of the odds ratio (SE), and associated p-value, by Stage and diagnosis, for the most significant SNP at the *PRSS1-PRSS2* locus and the *CLDN2* locus.

		***PRSS1-PRSS2* locus rs10273639 (T allele)**			***CLDN2* locus rs12688220 (T allele)**		
Dx	**Data**	**OR***	**se(OR)**	**p-value**	**OR**	**se(OR)**	**p-value**
CP	**Stage 1**	0.713	0.044	3.0×10^{-8}	1.612	0.081	2.4×10^{-21}
CP.RAP	**Combined**	0.734	0.030	2.0×10^{-14}	1.385	0.046	2.3×10^{-22}
CP	**Stage 2**	0.633	0.055	1.4×10^{-7}	1.336	0.092	2.6×10^{-5}
CP	**Combined**	0.683	0.034	1.7×10^{-14}	1.496	0.060	1.0×10^{-23}
CP.RAP	**Stage 2**	0.745	0.040	7.5×10^{-8}	1.238	0.056	2.3×10^{-6}

* The model used here includes covariates to control for the two leading eigenvectors for ancestry, as was done in the Plink analyses, but differs in its treatment of the minor allele count for the *CLDN2* locus, which resides on the X chromosome. In this case, Plink encodes the count of minor alleles in males as 0 and 1 and includes a sex effect; instead, following Clayton (2009; PMID: 19939292), we model the genotypes as 0 and 2 and do not include the sex effect, because this is a more powerful approach.

Supplementary Table 4. Distribution of rare, exonic risk variants in *PRSS1* by genotype at rs10273639 (C is risk allele for rs10273639). These results indicate that the N291 risk allele resides on the T haplotype of rs10273639 (non-risk), whereas the A16V and R122H alleles reside on the C haplotype (risk).

	All individuals			All cases (CP + RAP)		
Sequence variant	TT	CT	CC	TT	CT	CC
Wild Type	151	505	459	84	337	325
A16V		2	2		2	1
N29I	1	2		1	2	
R122H		8	8		8	8

Results are from NAPS2 samples, which were analyzed for rare variants over the duration of the NAPS2 study; the method of DNA analysis varied over time. Three rare variants – A16V, N29I and R122H of *PRSS1* (n=1112 with complete data on rare and common variant genotypes) – were evaluated by using a combination of methods, such as Surveyor (Transgenomic-Omaha, Nebraska) and Sequencing (Applied Biosystems-Carlsbad, California). The first batch of 950 samples were screened for exonic base-pair changes by Surveyor, and all positive and randomly selected negative samples were confirmed by Sequencing (Big Dye Terminator v3.1 Cycle Sequencing Kit cat# 4337456). All additional samples (n=665) were performed by Sequencing methods only. In brief, the Surveyor methodology (Surveyor Mutation Detection Kits cat #706020; http://www.transgenomic.com/pd/surveyor/Surveyor.asp) uses four steps: PCR amplification of DNA from wild-type and unknown sequence, followed by a hybridization step that forms hetero- and homo-duplexes, after which an enzyme that cuts at mismatched nucleotides is added, and finally DNA fragments are size-separated by gel electrophoresis.

We also validated genotypes of rs10273639 by independent TaqMan genotyping two SNPs in linkage disequibrium with it – rs2011216, in intron 1 of *PRSS1*, and rs6667, synonymous variant in exon 5 of *PRSS1* – in 1158 NAPS2 case and control subjects with complete SNP data. For all but 2 of the 1158 samples (99.83%), rs10273639 TaqMan genotypes were identical to those from OmniExpress. Alleles of the synonymous variant rs6667 were in perfect linkage disequilibrium with those at rs10273639, whereas alleles at rs2011216 showed a modest departure (data not shown). These results provide independent validation of rs10273639 genotypes, thereby confirming their association with risk for pancreatitis (Figure 1), and identify other SNPs that would produce genomewide association, including a synonymous SNP.

Supplementary Table 5. Normalized gene expression from pancreatic tissue for *PRSS1*, *PRSS2*, and *CLDN2*. Genotypes for rs10273639 (*PRSS1-PRSS2* locus) and rs12688220 (*CLDN2* locus) are given in the last two columns. "Study" refers to the source of each of 3 sets of samples analyzed, "ID" distinguishes subject sample, and NA means not available.

Study	ID	*PRSS1*	*PRSS2*	*CLDN2*	rs10273639	rs12688220
PITT	A	0.218	0.029	NA	TT	NA
PITT	B	0.218	0.090	NA	CT	NA
PITT	C	0.797	0.029	NA	CT	NA
PITT	D	0.877	0.593	NA	CT	NA
PITT	E	0.840	0.557	NA	CT	NA
PITT	F	1.657	1.319	NA	CC	NA
PITT	G	0.593	0.213	NA	CC	NA
PITT	H	1.459	1.470	NA	CC	NA
PITT	I	0.738	0.598	NA	CC	NA
PITT	J	0.914	0.670	NA	CC	NA
PSU	XBB 048	0.426	1.437	3.778	CT	TT
PSU	XBH 221	3.160	1.913	2.333	CT	CC
PSU	XBW 333	3.320	1.832	4.823	CC	TT
PSU	XEY 176	0.326	0.479	3.773	CT	CC
PSU	XFA 462	0.267	0.422	1.505	CT	TT
PSU	XFU 215	-0.238	0.296	2.690	TT	TT
PSU	XFY 110	-2.912	-2.752	-0.368	TT	TT
PSU	XGE 449	0.000	0.000	0.000	TT	CT
PSU	XGI 233	0.017	0.065	0.792	CC	TT
PSU	XHI 170	1.304	1.172	-4.853	CT	CC
PSU	XHN 379	-0.010	0.610	-1.307	TT	TT
PSU	XHN 423	-0.052	-1.155	-7.175	CC	TT
PSU	XIG 483	0.305	0.522	2.495	CT	CC
PSU	XIW 265	1.204	1.494	3.844	CT	TT
PSU	XJ 1285	3.524	2.351	4.950	CC	TT
PSU	XJD 339	-0.977	0.137	-6.026	CT	CC
PSU	XJD 439	0.041	0.449	2.159	NA	TT
PSU	XJD 447	2.479	2.729	-4.003	CT	TT
PSU	XJG 039	-0.134	-0.565	-6.244	CC	TT
PSU	XJG 404	1.255	1.183	2.412	CT	TT
PSU	XJH 462	2.220	0.885	2.149	CT	TT
PSU	XKB 098	-1.023	-1.355	-6.190	TT	CC
PSU	XKD 006	0.224	0.484	3.697	NA	TT
PSU	XKF 331	0.858	0.732	1.997	TT	TT
PSU	XKR 071	1.103	0.632	4.126	CT	TT
PSU	XKU 176	3.893	2.218	1.277	CT	CC
PSU	XL2098	2.698	-0.128	-3.505	CT	TT
PSU	XLE 240	4.046	3.805	-3.703	CT	CC
PSU	XLT 278	3.504	4.232	-2.162	CT	CC
PSU	YAC 312	0.399	1.769	3.485	TT	CC
PSU	YC 3158	-0.442	0.457	2.493	CT	TT

PSU	YC 4332	NA	1.623	3.303	TT	CC
PSU	YCC 461	2.403	0.979	1.647	CT	CT
PSU	YCH 191	-0.162	0.254	0.824	TT	CC
PSU	YCM 298	1.825	1.495	2.549	TT	TT
PSU	YCW 158	1.734	1.074	4.267	CC	TT
PSU	YCX 167	0.199	0.704	1.426	NA	TT
PSU	YCX 238	4.861	3.973	3.444	CT	TT
PSU	YDI 386	0.557	0.867	0.552	CT	TT
PSU	YDJ 130	0.077	0.434	2.690	CC	CT
PSU	YDX 455	0.069	0.941	5.515	CT	CC
PAGER	PA1318	-2.150	-2.380	NA	TT	CT
PAGER	PA1509	0.741	-1.936	-8.583	CC	CT
PAGER	PA1527	-0.510	-2.106	NA	CT	TT
PAGER	PA1553	-1.640	-2.622	NA	TT	TT
PAGER	PA1566	1.813	-1.743	-20.035	CT	TT
PAGER	PA1601	-0.436	-0.748	NA	CT	CC
PAGER	PA1615	-0.361	0.045	-0.855	TT	CT
PAGER	PA1641	-0.696	-1.011	-1.470	CT	CT
PAGER	PA1683	-0.128	0.035	2.255	CC	CT
PAGER	PA1690	0.000	0.000	0	CT	CT
PAGER	PA1701	-0.881	-0.775	-3.279	CC	TT
PAGER	PA1714	-0.316	-1.861	-2.993	TT	TT
PAGER	PA1728	-0.697	-1.056	-3.328	TT	CT
PAGER	PA1742	-0.798	-1.096	NA	CT	TT
PAGER	PA1761	-1.365	-2.529	2.756	TT	CC
PAGER	PA1796	-0.560	-2.208	NA	TT	CT
PAGER	PA1824	0.040	-1.270	0.298	TT	TT
PAGER	PA1855	-0.384	0.125	NA	TT	CT
PAGER	PA1880	-1.968	-1.490	NA	CT	CT
PAGER	PA1893	-2.351	-2.745	NA	CT	TT
PAGER	PA1900	-1.291	-0.568	NA	CT	CT
PAGER	PA1959	-0.896	-1.613	-4.529	CT	TT

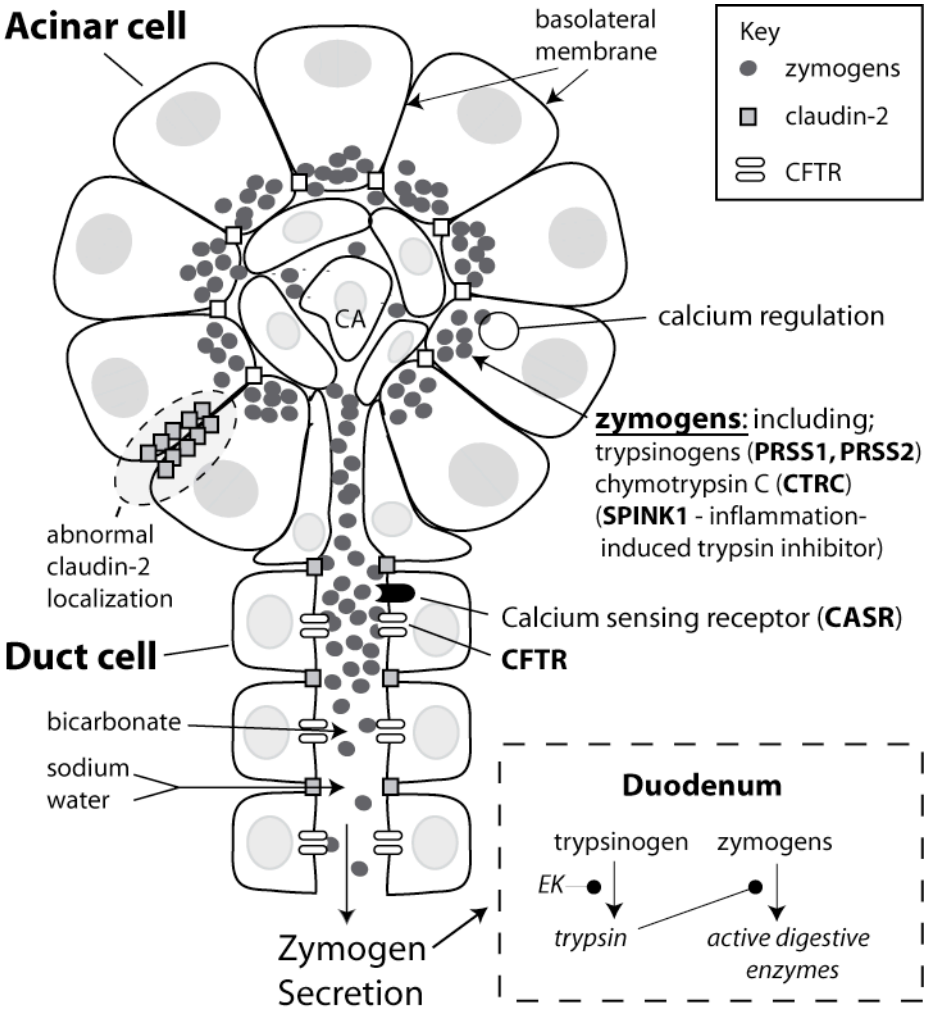
To determine whether *PRSS1* or *CLDN2* gene expression was associated with SNP genotypes (rs10273639 and rs12688220 respectively), we fit the normalized gene expression to counts of alleles. For rs10273639 (*PRSS1* locus), the count of alleles is 0, 1, or 2 C for risk alleles, whereas for rs12688220 (*CLDN2* locus), we used the allele encoding described in the footnote of Supplementary Table 2 (0/2 for males) with T being the risk allele. We fit the data for the three studies separately and then combined the results to form a single test statistic using a weighted sum of z-statistics, with the weights determined by sample size.

For *PRSS1* expression, the 3 slopes were 0.414 ± 0.1809 (± se); 0.756 ± 0.4094; and 0.301 ± 0.2904, with z-values 2.29, 1.85, and 1.04, which yields an overall p-value of 0.0099. (We removed 2 outliers, an observation with value < -4 from PAGER and one from PSU with value > 9.) For *CLDN2* expression, only 2 sample sets were characterized; the 2 slopes were -0.636 ± 0.5943 and 0.205 ± 1.4804, and z-values were -0.651 and 0.022, which yields an overall p-value of 0.32. (We removed one outlier from PAGER with value < -15.)

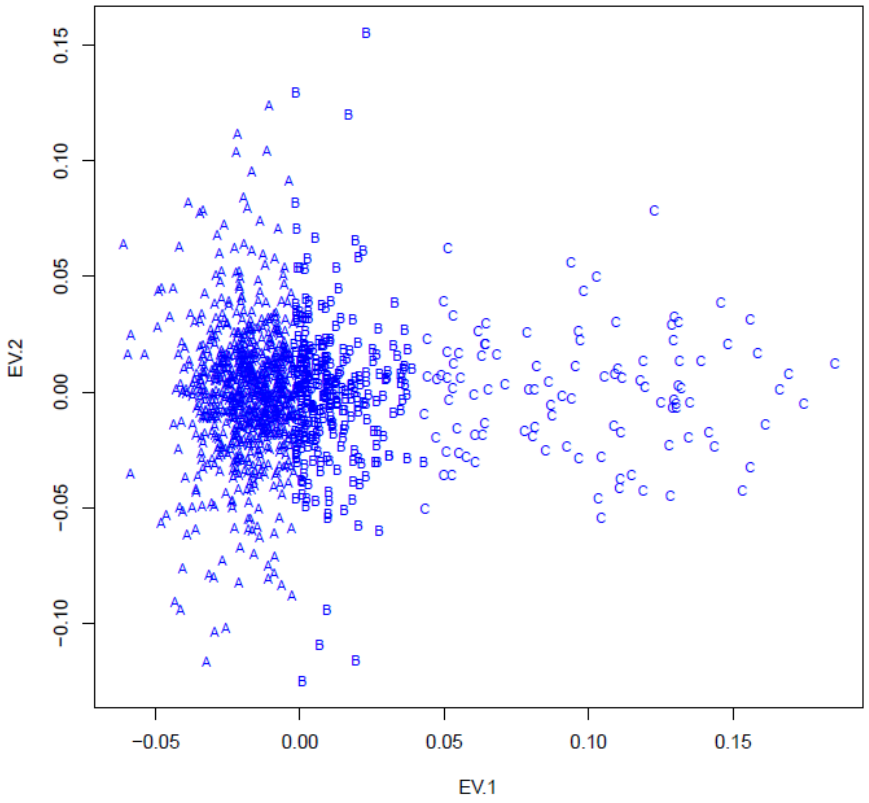
Supplementary Table 6. Histology and anti-claudin-2 staining of chronic pancreatitis tissue in GWAS cases. The samples are divided by rs12688220 genotype (1-2, low risk; = 7-12 high risk).

Claudin-2 staining		Intact acinar Cells		Atrophic acinar cells		Terminal ducts		Intralobular ducts		Islets
	Histology	Cytoplasm (granular)	membrane	Cytoplasm (granular)	membrane	Cytoplasm (granular)	membrane	Cytoplasm (granular)	membrane	
	Low-risk genotypes									
1	increased fibrosis, some residual lobules, some atrophy, chronic inflammation with lymphoid aggregates	++	-	++	-	+++	-	+++	-	+
2	residual acini, increased fibrosis, mild chronic inflammation	++	+	++	-	++	-	+++	-	++
3	residual acini with mild chronic inflammation; multifocal acute inflammation (?type 2 AIP?)	++	+	++	-	+++	-	+++	-	NP
4	atrophic, fibrotic, chronic inflammation	++		+++	-	+++	-	NP		NP
5	lots of fibrosis, mild chronic inflammation, some residual intact and atrophic lobules	++	+	++	-	+++	-	+++	-/+	++
6	intact lobules, increased fibrosis, mild chronic inflammation	++	-	++	-	++	-	+++	-/+	++
	High-risk genotype									
7	mostly atrophic with fibrosis and chronic inflammation	+	+++	-	P +++	++/+++	-	++/+++	f	++
8	increased fibrosis, some atrophy	-	+++	+	++++	+/++	-	na		NP
9	patchy chronic pancreatitis; mostly preserved	++	P ++	++	-	+++	-	+++	-	++
10	increased fibrosis, atrophy, rare residual lobule	++	P ++	++	-	+++	-	+++	-	+++
11	increased fibrosis, atrophy, rare residual lobule	++	P ++	++	-	++	-	++	-	++
12	increased fibrosis, atrophy, rare residual lobule	++	P ++	++	-	++	-	++	-	++

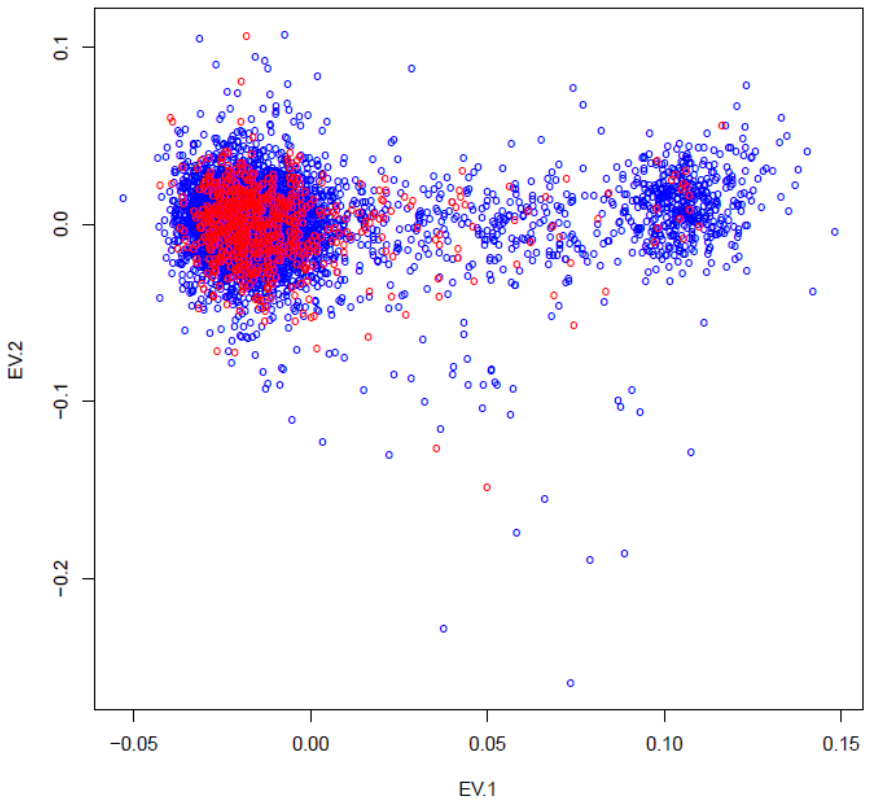
Staining intensity scale of – to ++++ (-, negative; +, weak; ++, moderate; +++, strong); f, focal; P, patchy; NP, structure was not present in the tissue section. Grey highlighted section identifies the primary feature that differentiates claudin-2 staining based on *CLDN2* genotype.



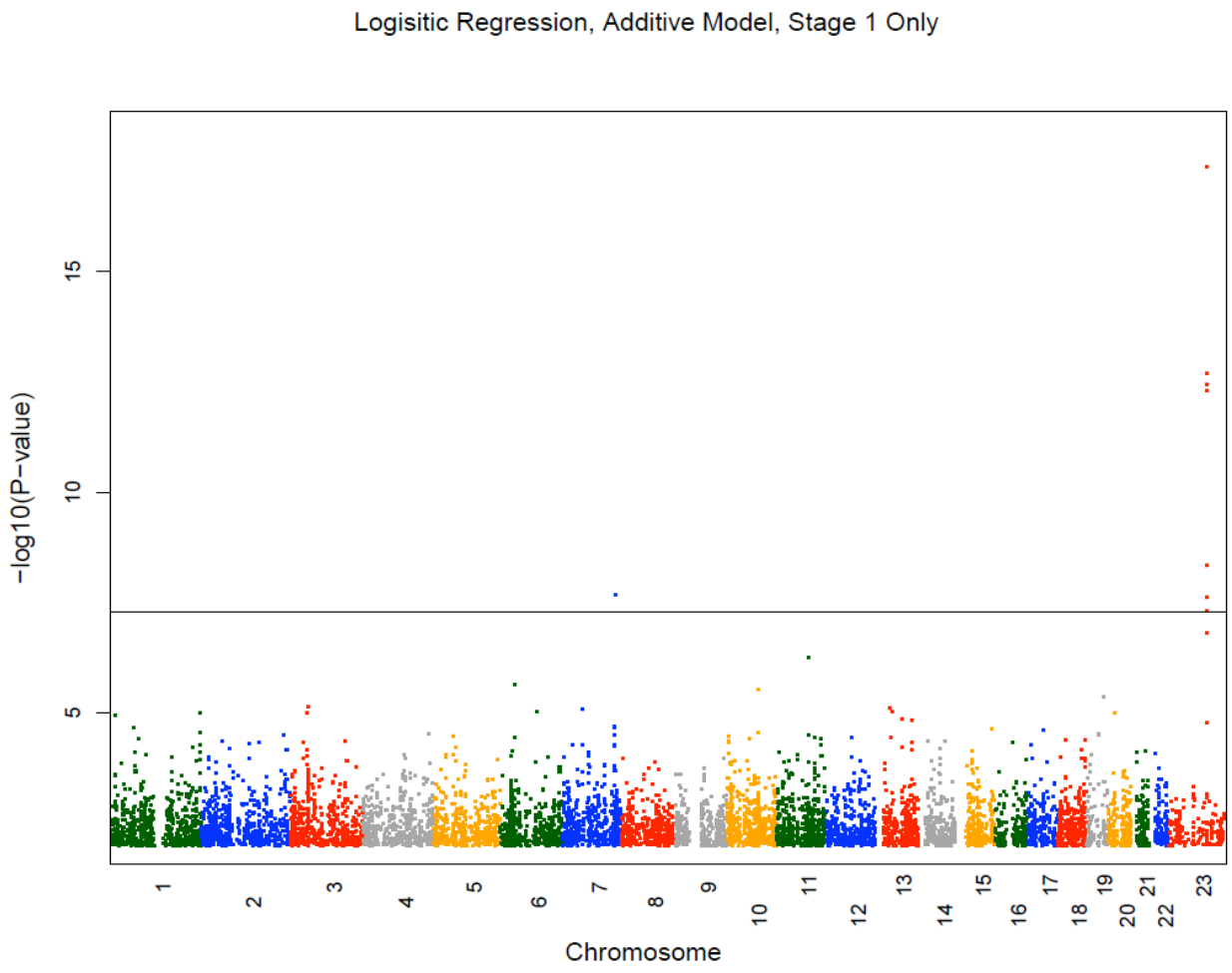
Supplementary Figure 1. Model of a pancreatic acinus illustrating pancreatic zymogen secretion from the acinar cells into the ducts, the relationship between acinar and duct cells, and the secretory pathway of the zymogens (digestive pro-enzymes), including trypsinogen to the duodenum, where they are activated. At least five genes have been identified as risk factors for chronic pancreatitis (CP) and replicated in multiple populations[2,3]. Gain-of-function mutations in the cationic trypsinogen gene (*PRSS1*) (e.g., p.N29I, p.R122H) cause hereditary pancreatitis[8], an autosomal dominant form of CP that begins with RAP and progresses to CP. Cationic and anionic trypsin are the master enzymes of the pancreas in that they regulate activation of the other pancreatic digestive zymogens following their own activation by enterokinase (EK) in the duodenum. Premature trypsin activation in the acinar cell or duct leads to zymogen activation, injury, and pancreatitis. Failure to inactivate trypsin directly (*SPINK1* mutations, *CTRC* mutations); to control calcium concentrations in the acinar cell (calcium regulation) or duct (*CASR* mutations) and thus enhance trypsin activation and survival; or to flush trypsin out of the pancreatic ducts (*CFTR* mutations) all increase the risk of developing pancreatitis. CA = centroacinar cells, a type of duct cell. Claudin-2 is normally expressed between duct cells, but abnormal localization may occur in association with a *CLDN2* risk allele (described in the text) during pancreatic inflammation.



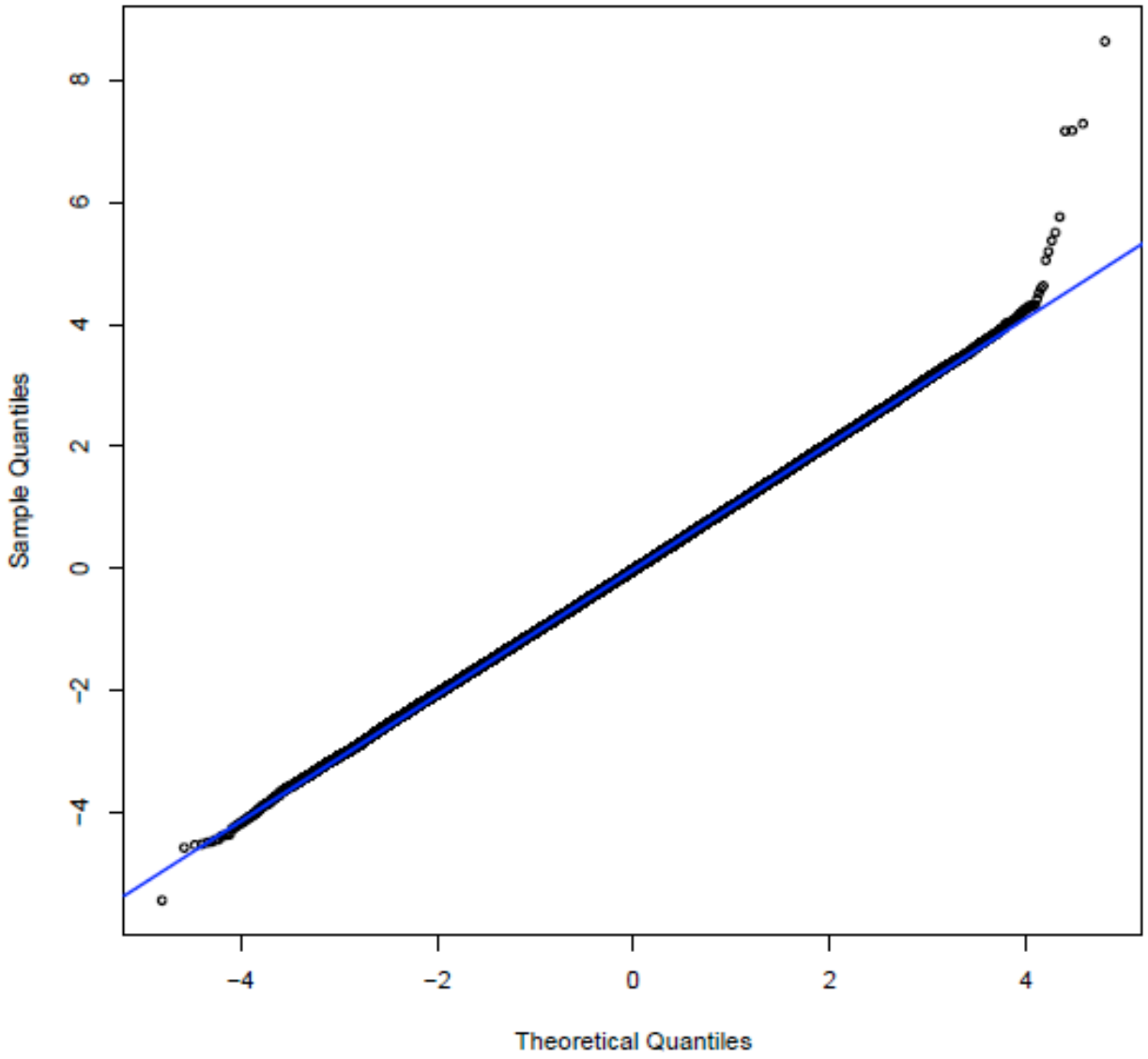
Supplementary Figure 2. Results from genetic ancestry analysis for chronic pancreatitis cases. dacGem (http://arxiv.org/abs/1104.1162) was used to convert multi-locus genotypes to ancestry dimension, expressed as eigenvectors. The decomposition resulted in a single significant dimension representing ancestry. Letters indicate ancestrally homogeneous clusters; of these clusters A and B are taken to represent a relatively homogeneous European ancestry. Note there was only one significant dimension of ancestry identified, although two are shown here.



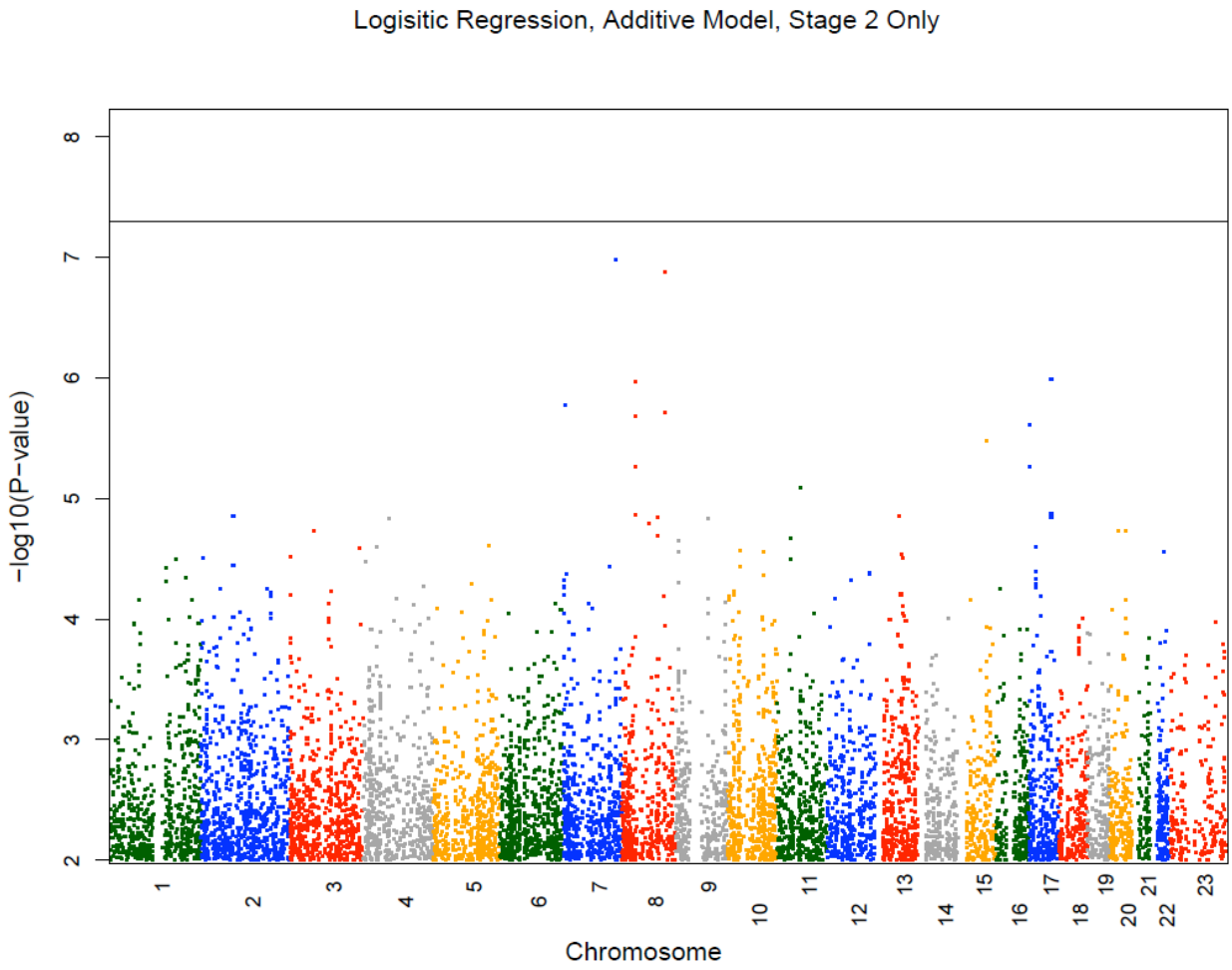
Supplementary Figure 3. Results from genetic ancestry analysis for chronic pancreatitis cases and ADGC controls. dacGem (http://arxiv.org/abs/1104.1162) was used to convert multi-locus genotypes to ancestry dimension, expressed as eigenvectors. The decomposition resulted in a single significant dimension representing ancestry. Red symbols are cases, blue symbols are controls. Note there was only one significant dimension of ancestry identified, although two are shown here.



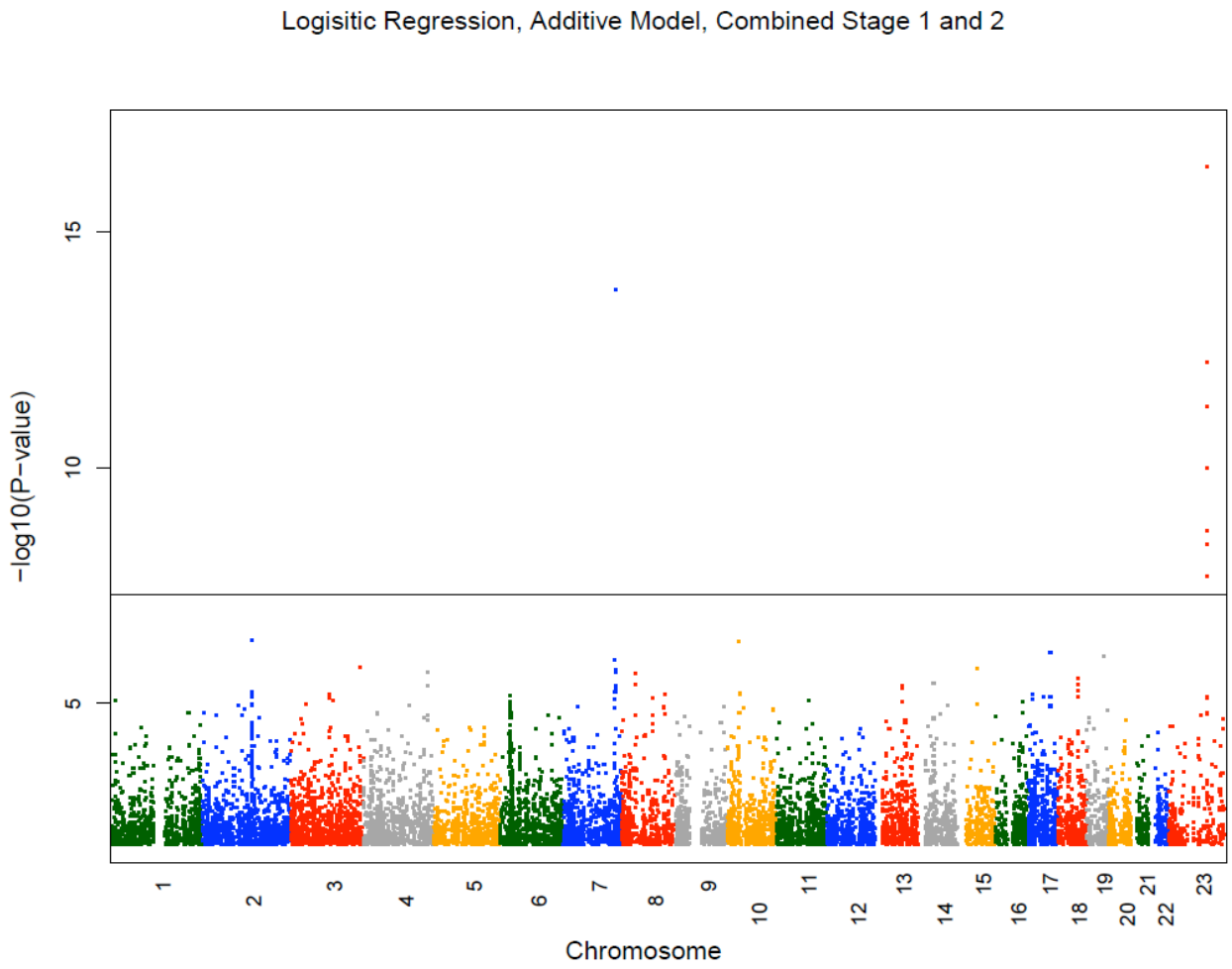
Supplementary Figure 4. Manhattan plot showing the negative log (base 10) of the p-value for association of SNP genotype with affection status (chronic pancreatitis versus control) for all SNPs passing quality control filters in Stage 1. Note the horizontal line at $5x10^{-8}$, which denotes an accepted significance threshold for genome-wide association. SNPs in interval Xq23.3 (*CLDN2* locus) cross this threshold, as does a SNPs in 7q34 (*PRSS1*-PRSS2 locus).



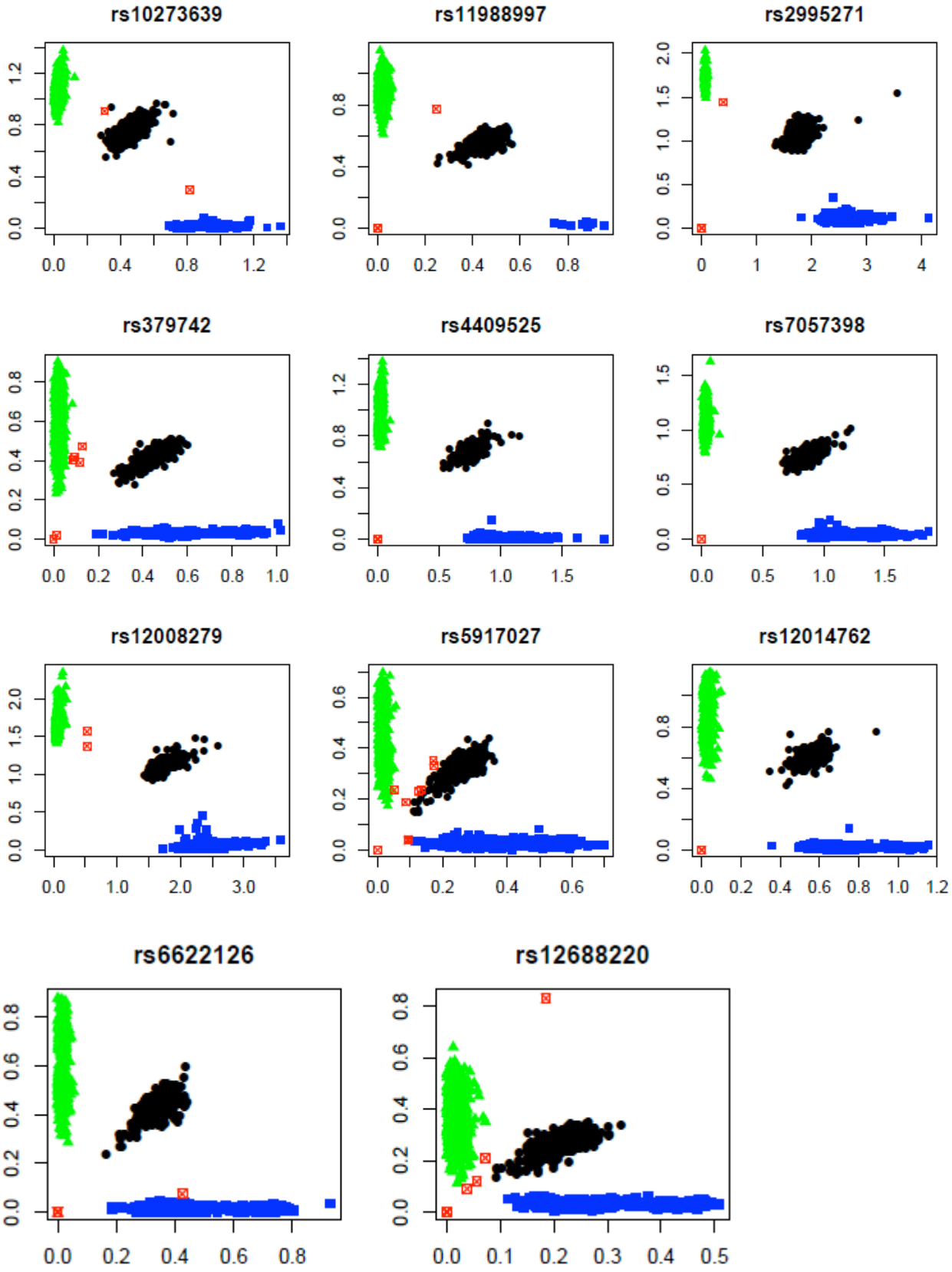
Supplementary Figure 5. Quantile-Quantile or Q-Q plot of association test statistics. Genomic inflation factor λ=1.06. If there were no difference between the observed distribution and that expected under the null hypothesis (no association), all points would fall on the line (X=Y).



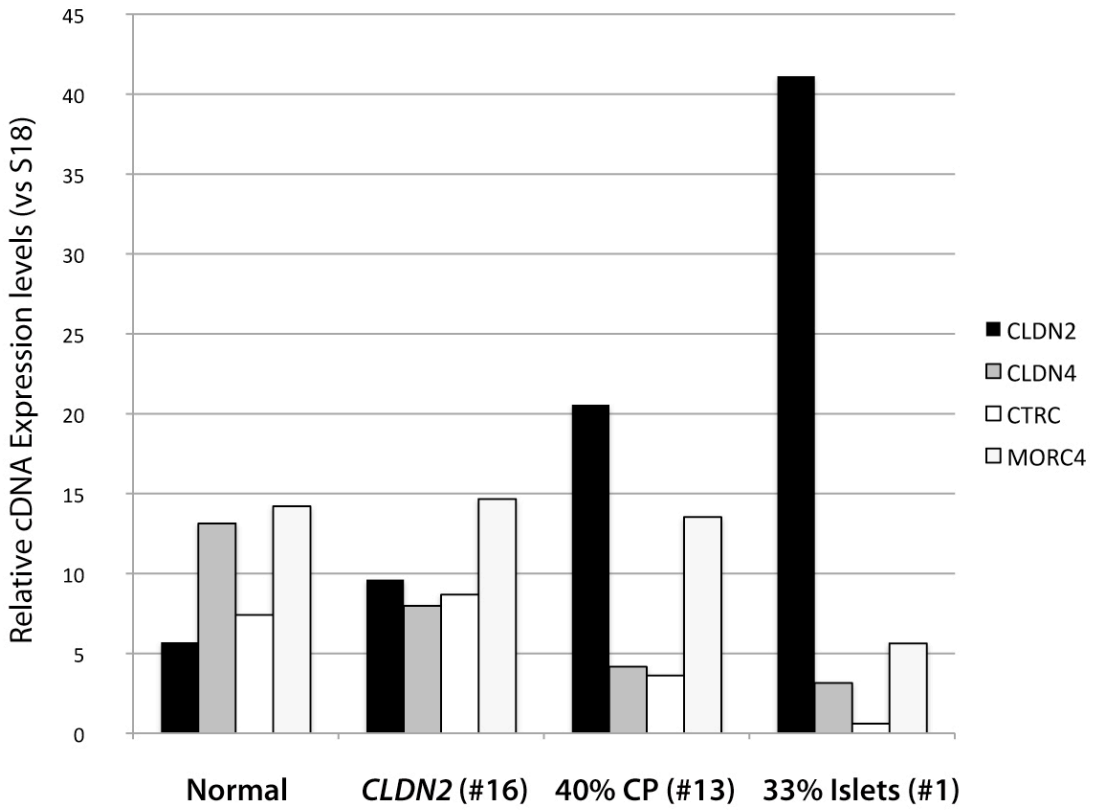
Supplementary Figure 6. Manhattan plot showing the negative log (base 10) of the p-value for association of SNP genotype with affection status (chronic pancreatitis and recurrent acute pancreatitis versus control) for all SNPs passing quality control filters in Stage 2. Note the horizontal line at $5x10^{-8}$, which denotes an accepted significance threshold for genome-wide association. Genomic inflation factor λ=1.09.



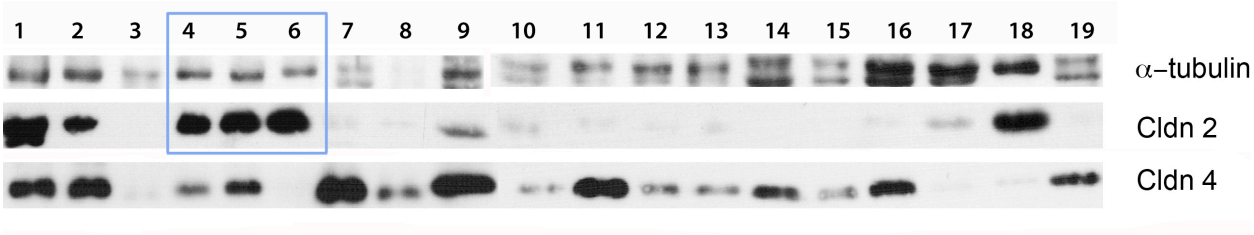
Supplementary Figure 7. Manhattan plot showing the negative log (base 10) of the p-value for association of SNP genotype with affection status (chronic pancreatitis and recurrent acute pancreatitis versus control) for all SNPs passing quality control filters in the combined data from Stages 1 and 2. Note the horizontal line at $5x10^{-8}$, which denotes an accepted significance threshold for genome-wide association. SNPs in interval Xq23.3 (*CLDN2* locus) cross this threshold, as does a SNP in 7q34 (*PRSS1-PRSS2* locus). Genomic inflation factor λ=1.13.


rs10273639
rs11988997
rs2995271
rs379742
rs4409525
rs7057398
rs12008279
rs5917027
rs12014762
rs6622126
rs12688220

Supplementary Figure 8. SNP genotype clusters for SNPs having p-values < 5×10^{-7} (as calculated by Plink) for the combined Stage 1 and 2 data. Green and blue indicate called homozygotes, black called heterozygotes, and red is non-called (missing) genotypes. Only SNPs having acceptable quality for genotype calls are shown. Note that rs10273639 (*PRSS1-PRSS2* locus) and rs12688220 (*CLDN2* locus) are the most highly associated SNPs for their respective loci.



Supplementary Figure 9. Relative expression of *CLDN2*, *CLDN4*, *CTRC*, and *MORC4* in control tissues. Surgical waste from grossly normal pancreas was collected and divided for histological, protein, and RNA studies. Relative expression of cDNA was analyzed with quantitative PCR using S18 RNA for normalization. Normal, average expression of 6 control samples with normal-appearing histologoy; *CLDN2* genotyped as homozygous for rs12688220 A (control sample 16); 40% chronic pancreatitis (CP) on histological evaluation (control sample 13); and 33% islets on histological evaluation (control sample 1). Expression of *CLDN2* appears to increase and *CLDN4* to decrease with the high-risk *CLDN2* genotype and in inflammation and islets, as expected[10,21,22]. *CTRC* is used as a marker of acinar cell gene expression, and *MORC4* is another candidate gene within the *CLDN2* locus. *MORC4* expression does not correlate with *CLDN2* expression or tissue histology.



Sample	Histology	Cldn-2	Cldn-4
Low risk (C/CC)			
1 female	33% acini/duct, 33% islets, 33% fat no inflammation	++++	+++
2 male	40% atrophic acini/duct, 33% fibrosis, 25% fat, 2% islets no inflammation	++	++++
5 male	25% acini/duct, 5% islets; 70% fibrosis focal chronic inflammation	++++	++
8 male	90% acini/duct, 5% islets, 5% fibrosis no inflammation	-	++
11 female	95% acini/duct, 2% islets,3% fat & fibrosis no inflammation	-	+++
17 male	85% acini/duct, 5% islets, 5% fat, 5% fibrosis, no inflammation	+	-
18 female	65% partially atrophic acini/duct, 5% fibrosis, 30% fat & fibrosis mild to moderate chronic inflammation	++++	+
19 male	all acini/duct, trace islets fat & fibrosis; no iInflammation	-	++
Low risk (CT)			
4 female	5% atrophic acini/duct, 45% fibrosis, 50% fat mild chronic inflammation	+++	+
7 female	90% acini/duct, 5% islets, 5% fibrosis no inflammation	-	++++
9 female	70% acini/duct 5% islets, 5% fibrosis 20% fibroadipose tissue no inflammation	+	++++
10 female	All cancer	+	+
12 female	67% acini/duct, 33% fat no inflammation	-	++
13 female	60% acini/duct, 40% fibrosis mild Chronic inflammation	-	++
15 female	all acini/duct, rare fat, fibrosis and islets no inflammation	-	+
High risk (T, TT)			

3	male	75% acini/duct, 20% fibrosis, 5% islets; no inflammation	-	-
6	female	80% atrophic acini/duct, 15% fibrosis, 5% islets mild chronic inflammation	++++	-
14	male	85% acini/duct, 5% fibrosis and islets, 10% fat no inflammation	-	++
16	male	90% acini/duct, 5% fat, 5% fibrosis and islets no Inflammation	-	+++

Supplementary Figure 10. Control pancreatic tissue Western blot. Effect of histology and rs12688220 genotype on claudin-2 expression. **Top panel**. Western blot of 19 control samples for claudin-2, α-tubulin (loading control), and claudin-4 (comparison molecule). Blue box outlines the samples used in the text (Figure 3A). Claudin-2 is up-regulated and claudin-4 down-regulated by inflammation in gallbladder[10]; this general pattern is confirmed in the pancreas. Claudin-2 is also expressed in islets, which were abundant in sample 1. **Bottom panel**. Details of the histological evaluation of the tissue adjacent to the sample used in the Western blot, organized by rs12688220 genotype. Cldn-2 and Cldn-4 are qualitative indicators of the density of staining in the top panel, to assist in review. The pattern is consistent with claudin-2 immunohistochemistry, which demonstrates strong up-regulation during inflammation (Figure 2 B non-inflamed versus C-E inflamed).

References

1. Szmola, R. & Sahin-Toth, M. Chymotrypsin C (caldecrin) promotes degradation of human cationic trypsin: identity with Rinderknecht's enzyme Y. *Proc Natl Acad Sci U S A* **104**, 11227-32 (2007).

2. Rosendahl, J. *et al.* Chymotrypsin C (CTRC) variants that diminish activity or secretion are associated with chronic pancreatitis. *Nat Genet* **40**, 78-82 (2008).

3. Lasson, A., Borgstrom, A. & Ohlsson, K. Elevated pancreatic secretory trypsin inhibitor levels during severe inflammatory disease, renal insufficiency, and after various surgical procedures. *Scand J Gastroenterol* **21**, 1275-80 (1986).

4. Ogawa, M. Pancreatic secretory trypsin inhibitor as an acute phase reactant. *Clin Biochem* **21**, 19-25 (1988).

5. Schneider, A. *et al.* Combined Bicarbonate Conductance-Impairing Variants in CFTR and SPINK1 Variants Are Associated With Chronic Pancreatitis in Patients Without Cystic Fibrosis. *Gastroenterology* **140**, 162-71 (2011).

6. Racz, G.Z. *et al.* Extracellular calcium sensing receptor in human pancreatic cells. *Gut* **51**, 705-11 (2002).

7. Dawra, R. *et al.* Intra-acinar trypsinogen activation mediates early stages of pancreatic injury but not inflammation in mice with acute pancreatitis. *Gastroenterology* **141**, 2210-2217 e2 (2011).

8. Amasheh, S. *et al.* Claudin-2 expression induces cation-selective channels in tight junctions of epithelial cells. *Journal of cell science* **115**, 4969-76 (2002).

9. Angelow, S., Ahlstrom, R. & Yu, A.S. Biology of claudins. *American journal of physiology. Renal physiology* **295**, F867-76 (2008).

10. Laurila, J.J. *et al.* Tight junction proteins in gallbladder epithelium: different expression in acute acalculous and calculous cholecystitis. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* **55**, 567-73 (2007).

11. Mankertz, J. *et al.* TNFalpha up-regulates claudin-2 expression in epithelial HT-29/B6 cells via phosphatidylinositol-3-kinase signaling. *Cell and tissue research* **336**, 67-77 (2009).

12. Suzuki, T., Yoshinaga, N. & Tanabe, S. Interleukin-6 (IL-6) regulates claudin-2 expression and tight junction permeability in intestinal epithelium. *The Journal of biological chemistry* **286**, 31263-71 (2011).

13. Schmid-Kotsas, A. *et al.* Lipopolysaccharide-activated macrophages stimulate the synthesis of collagen type I and C-fibronectin in cultured pancreatic stellate cells. *The American journal of pathology* **155**, 1749-58 (1999).

14. Van den Bossche, J. *et al.* Claudin-1, claudin-2 and claudin-11 genes differentially associate with distinct types of anti-inflammatory macrophages in vitro and with parasite- and tumour-elicited macrophages in vivo. *Scandinavian journal of immunology* **75**, 588-98 (2012).

15. Whitcomb, D.C. *et al.* Multicenter approach to recurrent acute and chronic pancreatitis in the United States: the North American Pancreatitis Study 2 (NAPS2). *Pancreatology* **8**, 520-31 (2008).

16. Whitcomb, D.C. *et al.* Angiopoietin-2, a Regulator of Vascular Permeability in Inflammation, Is Associated With Persistent Organ Failure in Patients With Acute Pancreatitis From the United States and Germany. *Am J Gastroenterol* **105**, 2287-92 (2010).

17. Brand, R.E. *et al.* Serum Biomarker Panels for the Detection of Pancreatic Cancer. *Clin Cancer Res* **17**, 805-816 (2011).

18. Applebaum-Shapiro, S.E. *et al.* Hereditary Pancreatitis in North America: The Pittsburgh - Midwest Multi-Center Pancreatic Study Group Study. *Pancreatology* **1**, 439-443 (2001).

19. Naj, A.C. *et al.* Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nature genetics* **43**, 436-41 (2011).

20. Hamza, T.H. *et al.* Genome-wide gene-environment study identifies glutamate receptor gene GRIN2A as a Parkinson's disease modifier gene via interaction with coffee. *PLoS genetics* **7**, e1002237 (2011).

21. Aung, P.P. *et al.* Differential expression of claudin-2 in normal human tissues and gastrointestinal carcinomas. *Virchows Archiv : an international journal of pathology* **448**, 428-34 (2006).

22. Lee, J.H. *et al.* Immunohistochemical analysis of claudin expression in pancreatic cystic tumors. *Oncology reports* **25**, 971-8 (2011).